DOCUMENT RESUME

ED 412 915 IR 018 576

TITLE	Internet-Accessible Scholarly Resources for the Humanities and Social Sciences.
INSTITUTION	American Council of Learned Societies, New York, NY.
PUB DATE	1997-02-00
NOTE	25p.; Second Series.
PUB TYPE	Collected Works - Serials (022)
JOURNAL CIT	ACLS Newsletter; v4 n4 Feb 1997
EDRS PRICE	EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Computer Networks; *Educational Researchers; Electronic Publishing; Electronic Text; Futures (of Society); *Internet; Newsletters; Online Searching; *Reference Materials; *Research Tools; Search Strategies; Students; User Needs (Information)
IDENTIFIERS	*American Council of Learned Societies; Digital Imagery; Technology Integration

ABSTRACT

This newsletter focuses on the presentations of a program session on Internet-accessible scholarly resources, held at the 1996 ACLS Annual Meeting. Articles in the newsletter include: "Building the Scene: Words, Images, Data, and Beyond" (David Green); "Electronic Texts: The Promise and the Reality" (Susan Hockey); "Images on the Internet: Issues and Opportunities" (Jennifer Trant); "The World Wide Web as a Resource for Scholars and Students" (Richard C. Rockwell); "The American Arts and Letters Network (AALN)" (Charles Henry); "The National Initiative for a Networked Cultural Heritage (NINCH)" (David Green); and "Because It's Time: A Commentary on the Program Session" (Willard McCarty). An alphabetical list of 42 URLs (uniform resource locators) for Web resources mentioned in the issue is included. The newsletter concludes with an excerpt from an obituary for Paul Evan Peters (1947-1996), a partner with the ACLS in the establishment of the National Initiative for a Networked Cultural Heritage. (SWC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 412 915

ACLS

Newsletter

Volume 4, No. 4 (second series), February 1997

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

Candace Frede

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Internet-Accessible Scholarly Resources for the Humanities and Social Sciences

Building the Scene: Words, Images, Data, and Beyond

by David Green
Executive Director, National Initiative for a Networked Cultural Heritage

Poets and philosophers have been thinking and dreaming about automata . . . at least since Homer. It is only in the last few years that computers have become sophisticated enough even to begin to handle our data and so be worthy of our attention.

—*Willard McCarty*

This timely panel, offered as a survey of what is currently available online for scholars, opened with presentations on the basic components of scholarly research: Susan Hockey on electronic texts, Jennifer Trant on digital images, and Richard Rockwell on digital data. It then built towards a synthesis with Charles Henry presenting the American Arts and Letters Network, an evolving online guide to exemplary digital resources in the humanities and, finally, Willard McCarty, as commentator, using what he heard to build a "typology of activities for scholarly Internet resources," a provisional blueprint for community "next steps."

Panelists, however, quickly surpassed their initial task of describing the immediate landscape by offering a continual critique of the medium and a call to the community to organize itself to actively influence the future shape of the Internet and its tools. They collectively understood that the Internet was passing through its adolescent, pioneering stage and, assessing its true powers and potential, was beginning to develop some really useful strategies for the future. The World Wide Web, essentially a display and linking medium, is currently the most popular form of the Internet. While it presents text, images, and sound well and hyperlinks between distributed

Continued on next page

Contents:

This issue focuses on the presentations of a program session on Internet-accessible scholarly resources held at the 1996 ACLS Annual Meeting.

At the Annual Meeting, the ACLS introduced its Web site: **http://www.acls.org**. We invite you to visit the site for online versions of this and other ACLS publications.

Building the Scene: Words, Images, Data, and Beyond 1
by David Green
Electronic Texts: The Promise and the Reality 3
by Susan Hockey
Images on the Internet: Issues and Opportunities 6
by Jennifer Trant
The World Wide Web as a Resource for Scholars and Students 9
by Richard C. Rockwell
The American Arts and Letters Network (AALN) 11
by Charles Henry
The National Initiative for a Networked Cultural Heritage (NINCH) 12
by David Green
Because It's Time: A Commentary on the Program Session 14
by Willard McCarty
URLs for Resources Mentioned in this Issue 23
Paul Evan Peters (1947-1996) 24

BEST COPY AVAILABLE

Building the Scene

Continued from preceding page

resources brilliantly, it is abysmal as a tool to efficiently find material and totally incompetent in managing text or enabling its sophisticated use or understanding.

Texts

Susan Hockey, founding director of the seminal Center for Electronic Texts in the Humanities, dated the birth of humanities computing back half a century to Father Busa's 1948 electronic concordancing of the works of Thomas Aquinas. Hockey agrees with Father Busa's own recent judgment that in the interim there has been surprisingly little intellectual progress in the development and sophisticated use of electronic texts.

Hockey's adept guide to the state of the art of creating and using e-texts (including publishing multi-dimensional scholarly editions) pointed out the kinds of primary source documents now available in electronic form; the ways these digital texts can be used; how to best create an electronic text oneself; how the coding or "markup" of the structure of e-texts can make them significantly more useful for scholars; how electronic texts are delivered—on CD/ROM or via the Internet; and the under-used opportunities to use hypertext linking to create new forms of publication, often better suited to scholarly inquiry than traditional linear sequencing.

From her vantage point, Hockey was able to clearly delineate the landscape of what could be done using the tools of today. However, this survey led to the major theme of her presentation: a call for the community to be more pro-active in defining the tools and materials it needs for the future. The real challenge is not simply to make today's scholarly practices more effortless by enabling easier access to material (though that is a major and worthy task), but rather to pursue the potential inherent in the medium to examine, analyze, manipulate, critique and create texts in new ways.

This challenge operates on several levels: first to be aware of the immense added value given to text that is encoded or "marked up" using Standard Generalized Markup Language (SGML). SGML, through its tagging system, provides a solid basis for describing the structural characteristics of a text (or object) and its content, irrespective of hardware or software platforms. It thus makes the text easier to query and more productive to analysis. Encoding in SGML requires a thorough analysis of the intellectual issues involved in creating an electronic object: what it is and how it may be used. SGML-encoded text can only be read in a dumbed-down version through the HyperText Markup Language (HTML) of the Web as we know it today. So there is urgent need for a more sophisticated post-Web Internet medium.

Most sophisticated electronic texts are in fact not yet available on the Internet. With these texts, mostly delivered by CD-ROM, there is immense work to be done, with software developers but also among scholars, to develop software programs that will allow them to be used more flexibly and powerfully than they are today.

Images

Compared to the long history of electronic text, digital images have a less extensive history, although the World Wide Web has made imagery much more omnipresent on the Internet. Some argue that, regardless of the Web, images are being used increasingly today by many humanities scholars, not just art historians and not just for simple illustration. One of Jennifer Trant's interesting lines of thought concerned the reconceptualization of visual information sources today. While indicating challenges ahead for a medium less well developed than text, Trant focused more on charting the essential sources for images available for scholars in electronic form today.

The processes and issues of transforming text and image into digital forms are different, if related. Text can either be laboriously typed into computer memory or scanned using Optical Character Recognition (OCR) software (that converts the page images of text into manipulable words), depending on the accuracy needed. Text pages can also be simply scanned into a computer

Continued on page 20

ACLS

American Council of Learned Societies
228 East 45th Street
New York, NY 10017-3398
212 697-1505
http://www.acls.org

The American Council of Learned Societies was founded in 1919. It is a private nonprofit federation of 58 national scholarly organizations. The purpose of the Council, as set forth in its constitution, is "the advancement of humanistic studies and the maintenance and strengthening of relations among the national societies devoted to such studies." The Council, consisting of a 15-member Board of Directors and one Delegate from each of the constituent societies, meets annually.

The **Newsletter** welcomes communications from readers. Please address correspondence to Candace Frede, Director of Publications, American Council of Learned Societies, 228 East 45th Street, New York, NY 10017-3398; candace@acls.org.

Electronic Texts: The Promise and The Reality

by Susan Hockey
Director, Center for Electronic Texts in the Humanities

With the current level of interest in the Internet, World Wide Web, and digital libraries, electronic texts and the technology needed to use them are now in the center of the scholarly arena. But their usage in humanities scholarship is not new. It is now almost fifty years since the first electronic text project began in 1948 when Father Busa started to prepare electronic texts of St Thomas Aquinas and related authors for his monumental Index Thomisticus.[1] Since then a considerable body of expertise has developed among scholars who work with electronic texts, but so far less of this expertise than one might hope has found its way to most of the texts available on the Internet today.

Characteristics of Humanities Texts

This piece concentrates on primary source texts which are studied for many purposes and are the most varied in nature. The solutions to problems encountered in working with electronic primary source material can be applied just as well to simpler material such as journal articles and monographs. Humanities primary source texts may take the form of literary works, historical documents, manuscripts, papyri, inscriptions, coins, transcriptions of spoken texts, or dictionaries and they may be in any natural language. Much of this material is complex. It may include variant readings, variant spellings, marginal notes, annotation of various kinds, cancellations and interlineations as well as nonstandard characters. Much of the scholarship carried out on this material consists of interpretation and annotation, often involving fine detail as well as broad overviews.

Representing this material in electronic form in such a way that it can be used for traditional scholarly processes in the humanities poses significant intellectual and technological challenges, especially when, as is often the case, scholars differ in their opinions about the material.

Uses of Electronic Texts

Electronic texts can be used for many purposes in the humanities.[2] Content retrieval, where, for example, the computer is instructed to find all the documents about a particular topic, is only one application. Other scholarly uses include rapid retrieval of words and phrases for the study of lexis and syntax, and the production of concordances and word frequency lists on which analyses of vocabulary, style, disputed authorship, and poetic rhythm and meter can be based. Other projects have used concordances in the compilation of historical dictionaries, and for scholarly editions. These fairly mechanical processes are relatively well understood and they have been used intelligently by many projects, but they do have significant limitations. Words are seen only as a sequence of letters. There are no facilities for separation of homographs (e.g., "might" as an auxiliary verb and as a noun), or for morphological analysis (e.g., for inflected languages such as Latin and Greek), or for lemmatization (putting words under their dictionary headings), or for bringing together variant spellings of the same word.

Automatic concept-based searching, which is what most scholars really want to do, is still a long way off.

Markup

Markup makes explicit for computer processing things which are implicit for the human reader. We are of course most familiar with typographic markup, but this is ambiguous for most kinds of analysis of electronic texts.[3]

For example, italics can be used for titles, foreign words, emphasized words or stage directions. It would not be possible to use italic markup to make an index of titles or exclude stage directions from a word search.

Structural markup distinguishes titles, foreign words, emphasized words, stage directions and anything else. It can be used to refine searches and can also be mapped on to typographic markup for display or printing. HTML is a specific markup format which is interpreted by World Wide Web browsers to display text or provide links to graphics, sound or to other HTML files. HTML is a mixture of structural and typographic markup, but the structure which it uses is not rich enough for most kinds of searches or analyses of humanities texts.

Standard Generalized Markup Language (SGML) and the Text Encoding Initiative (TEI)

SGML is a kind of computer language which permits the definition of many different markup schemes.[4] HTML is in fact a simple SGML-based scheme. The TEI has defined an SGML-based markup scheme which can handle many different types of humanities texts. The TEI *Guidelines for the Encoding and Interchange of Electronic Text* were first published in 1994 after six years of work based on the experiences of an international community of humanities computing scholars.[5] The TEI scheme can be used in a very simple fashion, but it can also handle very complex texts and represent multiple and conflicting views within those texts. SGML makes the computer processing of texts easier because of the SGML Document Type Definition, a formal specification of the document structure which is used by any SGML-based computer program to derive information about what features the program can expect to find in the text. SGML files are important for the longevity of electronic texts since they are plain ASCII and can thus move very

Continued on next page

Texts

Continued from preceding page

easily from one computer system to another. They can also be processed by any SGML-aware program and so the user of SGML texts is not dependent on any one software vendor. SGML is now being widely adopted in industry as companies realize the cost of moving from one software system to another. Many academic publishers are now also adopting SGML. Notable among these is Chadwyck-Healey, who use a version of the TEI for the English Poetry Database, the Patrologia Latina, and other textbases.

Creating Electronic Texts

Creating an electronic text can be a very time-consuming process. Keyboarding is the best method to reach the levels of accuracy that humanities scholars would expect to find. Even on texts which are clearly printed and all in the same typeface, optical character recognition (OCR scanning) is rarely more than 99.9% accurate, that is one error per 1,000 characters or 10-12 lines.[6] OCR systems typically have problems with early printed books, manuscripts, newspapers, dictionaries and bibliographies (where there are many changes of typeface), and microfilm prints. Even if the letters in the text are recognized accurately by the OCR program, work still needs to be done to create an accurate and usable text for processing. This may entail inserting markup and deleting extraneous information such as page numbers, line numbers, and other marginal information. It is worth noting that most of the major humanities electronic text projects (Thesaurus Linguae Graecae, Women Writers Project, Perseus, Oxford English Dictionary, etc.) do not use OCR, but prefer to keyboard their text. OCR may be more appropriate for content-based document retrieval where a completely accurate text is less important. In this case OCR is used together with digital images of text. A search is performed on the uncorrected text but the result is delivered to the user as a digital image of the page where the match is found. This is the approach being used by the JSTOR project funded by the Andrew W. Mellon Foundation, which will initially consist of about 10 journal titles in the areas of economics and history and will initially contain approximately 750,000 journal page images.

Metadata: Data about Electronic Texts

Before the TEI there was no common methodology for describing and documenting electronic texts. Scholars who created electronic texts for their own purposes did not feel the need to prepare documentation since they knew what the material was in any case. The large European archives of electronic text which began in the 1960s mostly use simple database tools for keeping records of what they have got. A scholar looking for an electronic text of an author is particularly interested in what source text was used, what features are encoded in the text, and what changes have been made to the electronic text. None of these items of information fits well into library cataloging models such as MARC, which are also expensive to create and maintain. At the other extreme, the current version of HTML has very little provision for metadata, which is why the current WWW crawlers can perform only very crude searches. The TEI has developed a specification of an electronic text file header which provides all of this information and can also act as a worksheet for a full catalog record if that is needed. Since it is also in SGML the TEI header has the added advantage of being processable by the same software as the rest of the text. It can thus remain part of the text file rather than be separated from it. In an attempt to make some progress with metadata for WWW documents, in 1995 OCLC convened a meeting of experts who defined a set of core elements for the description of networked document-like objects, known as the "Dublin Core."[7] A further meeting in early April 1996 discussed ways of implementing this, possibly using SGML.

Electronic Text Products

Electronic texts are distributed either as raw text files which then require a third-party program, or they are packaged together with a retrieval program. Most publishers' electronic text CD-ROMs (Chadwyck-Healey English Poetry Database, OED, CETEDOC CD-ROM, etc) are packaged with proprietary software. It may take some time to learn how to use the software, but more experienced users also often find that this software does not answer some of the questions they want to ask. Libraries are also finding it expensive to support many different products. Raw text files need to be indexed for most applications. Otherwise searches can be very time-consuming. Indexing large textbases with Opentext's Pat program or similar tools is not a trivial task and the burden of deciding which options to choose in the indexing program and thus what words can be retrieved often falls on a computer systems person who may well not understand fully what scholars might want to do with the texts.

Electronic Texts on the Internet

The electronic texts which are available on the Internet today represent only a small portion of the total number of texts in existence. The World Wide Web's strength is that it can bring to the desktop material which

Texts

Continued from preceding page

would otherwise be inaccessible. However the Web is basically a delivery mechanism. It provides only limited means of manipulating and analyzing the texts. At some institutions the Web and HTML are used as a means of delivering the results of retrieval programs which operate on a richer markup scheme such as the TEI. The results of these searches are normally converted ("dumbed down") to HTML for display to the user, but they can also be delivered in SGML for further manipulation.[8] The Web's limited facilities for metadata are an added problem for the humanities. Current Web technology also encourages the fragmentation of material into short files which may not necessarily map well to the logical structure of a text.

It is worth noting that some of the major electronic textbases in the humanities are not yet on the Internet. These include the Thesaurus Linguae Graecae, Packard Humanities Institute Latin texts, and the Women Writers Project. This is partly because of licensing issues but also because of the inadequacies of current searching tools.

Exploiting the Electronic Medium for New Forms of Publication

For publication in electronic form, it is no longer necessary to write in a single linear sequence, as is the case for publication in print form. Hypertext permits links between associated, yet randomly distributed, items of information.[9] It has become very popular in the humanities largely because it can model the way scholars follow a thread from one source document to another. Yet many hypertexts, including much material on the World Wide Web, at present use electronic material as if it was in print form. Some groups of scholars are beginning to devise publications which can only exist in electronic form. Foremost among these in the humanities are editions where the electronic form permits multiple versions of the same text with links between transcriptions of the text, digital images of the manuscripts, and annotation of various kinds.[10]

The Model Editions Partnership (MEP; **http://mep.cla.sc.edu**) is a consortium of seven documentary editing projects, CETH and the TEI.[11] Funded by the NHPRC, the MEP is developing a set of models for electronic historical editions which will exploit the electronic medium for addressing questions of intellectual access to source documents and their context as well as maintaining current standards of scholarly editorial excellence. Scholars will be able to go from a clear reading text to the most conservative diplomatic transcription at a click of a mouse. An editor can embed variants and footnotes within the text so that scholars can view or suppress them at will. Instead of being limited to a single organizing principle like chronology, an edition can allow readers to dynamically organize documents into subsets relating to their own interests. Annotation can be linked to many documents and indexes prepared cumulatively. The MEP sees markup as the key to the functions it needs to provide and is creating a specialized version of the TEI encoding scheme. This will also enable print to be produced very easily if needed. In the first phase of the MEP, the three coordinators visited all seven partner projects to examine their methods of working. Building on the information acquired during these visits the MEP then drafted a prospectus for electronic historical editions, which was circulated widely among the documentary editing community for comment. The markup scheme is now being drafted and this will also be circulated and tested widely.

CETH is also participating in the Orlando project at the University of Alberta (**http://www.ualberta.ca/ORLANDO**). Orlando is creating an Integrated History of Women's Writing in the British Isles which will comprise four volumes and a chronology. SGML is being used throughout this project, including note-taking as the research for the project is carried out. The authors of the volumes will draw on a centralized SGML-encoded database of information as they write. The chronology can be produced almost automatically from this database and it is envisaged that there will be several other spin-off hypertext products all created by using the markup.

As the use of the Internet and electronic technology develops, we can expect to see many more projects like the MEP and Orlando, that is, projects which really use the medium in ways that otherwise would not be possible. At present much of the current Internet excitement is generated because of the easy access to material. As more material becomes available and more scholars use it, this excitement will begin to wear off. Usage will become commonplace, no different from how the library catalogue is used now. It is questionable how much of the current material on the Internet will be really usable and useful in this new environment. We must begin now to think what humanities scholars will want to do in the twenty-first century and plan the building of electronic resources to meet these needs.

Notes

1. Father Busa's remarks on the lack of intellectual progress since he began his project in 1948 are salutary reading for those embarking on ambitious projects. See "Half a Century of Literary Computing: Towards a 'New' Philology", in Reports of Colloquia at Tübingen, *Literary and Linguistic Computing* 7 (1992): 69-73.

Continued on page 22

Images on the Internet: Issues and Opportunities

by Jennifer Trant
Partner and Principal Consultant, Archives & Museum Informatics
(formerly Manager, Imaging Initiative, Getty Art History Information Program)

The potential for information networks to make visual information more readily available for research and education is vast and seductive. Distributing digital images on the Internet offers the potential to defy geography, providing access to national collections for students in remote areas. It can enable the scholarly study of physically separated objects, and open the eyes of new user communities to the possibilities of visual communications. However, this reconceptualization of visual information resources offers the greatest challenge. A new interest in the nature and power of visual information—partly driven by the need to satisfy the information collection and absorption tendencies of a young generation, steeped in a new vocabulary predicated upon quickly-paced aural and visual stimuli, and partly driven by the new communications possibilities of interactive multimedia and the World Wide Web—has created a rising demand for digital images.

This piece highlights some of the issues facing those involved with the management of information about works of cultural heritage, and offers a sense of the activities underway to address them. It focuses on images that document works of cultural heritage, purposefully excluding page-images of scanned text and vector or 3-D images (including scientific visualizations) as these data types, while similar, raise other specific questions. By focusing on those aspects of the information carried by images that make them intrinsically different than text, we can ask "What are the preconditions to using visual materials across disciplines?"

The issues fall broadly into three areas: 1) the construction of information resources; 2) the use of information about works of art; and 3) the implications for those involved in teaching with visual images.

Resource Creation

For distributed collections to work together on the network, they need to be constructed according to commonly known standards. Otherwise, it will not be possible for a scholarly user to assess the informational content of a digital image, and judge its fitness as evidence to support the research process. As yet, however, there are few standards for the capture, description, and distribution of images, and the quality of material available on the Internet ranges widely.

What is a Digital Image?

A digital image is a composed of a set of *pixels* (picture elements), similar to dots on a newspaper photograph or the tiles in a mosaic. Each pixel represents a portion of the image in a particular color, or shade of gray. Three major criteria contribute to quality in the creation of a digital image: 1) the number of pixels in a particular area, defined as *resolution*; 2) the number of possible colors for each pixel, defined as *dynamic range*; and 3) the method by which an image was converted to digital form, known as *scanning*.[1]

Image Quality

Once an image is converted into digital form, it is often compressed for storage; this process can either be *lossy*, wherein information is eliminated from the original digital image file, or *lossless*, wherein it is possible to reconstruct the original file, exactly as it was before compression. Together, these characteristics define how much information is recorded in a digital image, and how faithful a digital reproduction may be to an original.

Consistent approaches to digital conversion are essential to create a predictable resource of even quality. Without such procedures, the reliability of digital images is questionable, and their usability for scholarly research is uncertain. For example, color monitors vary greatly in their display of images; studying the palette of an artist using digital images is impossible if color management mechanisms are not in place. If networked digital images are to become a serious scholarly resource, we must find ways to assure their integrity and information content.

Image Description

Documentation is one part of such a strategy. Recent discussions within the image community have furthered our understanding of the need to record the processes of image capture, and to make this information available to the users of digital visual resources. A recent meeting of image specialists, sponsored by the Coalition for Networked Information (CNI) and OCLC, enumerated a number of technical characteristics of a digital image, and proposed a method for integrating them into a descriptive framework for documenting information about an image—what is referred to as *metadata*.[2]

These discussions are contributing to the definition of the "Dublin Core," a set of information categories to support information discovery on the Internet. If implemented, the Dublin Core could assist in improving the precision of searches of network indexes by enabling queries of such categories as "creator" and "title."

Resource Use

While sizeable collections of digital images are being created, several significant barriers to their use remain. These include difficulties in locating particular images, and once appropriate images have been discovered, in understanding and negotiating intellectual property rights agreements to use them. Consistent descriptions of images and the development of interdisciplinary, multimedia searching capabilities

Images

Continued from preceding page

will enable more efficient resource location. Common frameworks of rights, and shared procedures for negotiating usage of digital resources, will help bridge the copyright chasm. Several projects are now addressing these issues in a preliminary way.

Finding Images

Difficulties in finding images are familiar to those who study visual materials. For example, the works of a particular artist are often dispersed in public and private collections around the world. One of the great potentials for networked images is the possibility to integrate information about collections which exist in a number of places, making preliminary research a much less onerous task. While discussions such as that surrounding the "Dublin Core" categories for information discovery offer a longer term and more robust solution to information resource discovery across the Internet, a number of sites have grown up on the World Wide Web that offer help finding images. In the majority of cases, these are indexes which point to clusters of images made available by particular institutions.

ArtSource **(http://www.uky.edu/Artsource/artsourcehome.html)**, maintained at the University of Kentucky, is one of the oldest. As well as including general information about art and architecture, and art and architectural libraries, it includes a list of image collections, each accompanied by a brief summary of its content.

The Art History Department at the Australian National University, maintains a significant body of images at a site known as ArtServe **(http://rubens.anu.edu.au)**. ArtServe has grown from a site created particularly to support departmental teaching, where resources were grouped according to courses (for example, the site includes a set of images which document the history of European printmaking), to a more robust and searchable site, offering a range of images documenting many subject areas.

The Virtual Library Virtual Museum pages **(http://www.icom.org)** maintains a geographical index of museums with a presence on the Internet. As well as links to collections, this site includes links to many "virtual exhibitions." Museum exhibitions are frustratingly transitory, existing only for a short time in a particular place. More and more often these temporary assemblages of works of art are maintained online, making it possible to communicate a curatorial thesis beyond the closing date of a show.

There are also sites on the Web centered on the creation of works of art in digital form. fineArt forum **(http://www.msstate.edu/Fineart_Online/home.html)** is a site which focuses very much on the work of contemporary artists. As well as including a regular online magazine, *fineArt forum online*, this site includes an online gallery and links to resources including museums and galleries. ArtsWire **(http://www.artswire.org)** also offers online links to the works of visual artists and the issues that concern them.

There are also many idiosyncratic image sites on the World Wide Web at the moment. These range from the Christus Rex Web site **(http://www.christusrex.org)**, which offers one of the most comprehensive sets of images of the Vatican, including the restored Sistine Chapel ceiling (delivered along with a range of religious documents and papal encyclicals), to the WebMuseum, Paris **(http://sunsite.unc.edu/wm)**. Previously known as the "WebLouvre," this site offers a range of materials drawn from museum collections around the world, but has no formal institutional affiliation. Contrasting information from such informal sources with that offered by institutions about their own collections, in sites such as that of the National Museum of American Art **(http://www.nmaa.si.edu)**, raises significant questions about the source of visual information, its authenticity, and level of accuracy. We need to examine issues of authentication and quality assurance, so that we can be sure that the sites being used very actively by students are a valid research resource.

While each of these sites contains a wealth of visual information, it is not integrated in any real way. A researcher must still either browse lists of images available or conduct separate searches of each site, making the process a repetitive and often frustrating one. The user must also know which sites have visual images available, or be able to identify one of these starting points.

Implications

Interdisciplinary Use

Once images are located, there are still relatively few tools to support their actual use by scholars. Images are often delivered as part of a particular Web page. Little attention is paid to how an image might be reused in another context, and little information is often communicated with the image to enable further use. Some sites, however, are beginning to demonstrate the potential for the World Wide Web to integrate material of many media into thematically organized research resources.

The Center for Advanced Technology in the Humanities at the University of Virginia has begun to examine the integration of visual and textual material in a site devoted to "The Complete Writings and Pictures of Dante Gabriel Rossetti" **(http://jefferson.village.virginia.edu/rossetti/**

Continued on next page

Images

Continued from preceding page

rossetti.html). Developed by Jerome McGann and billed as a "Hypermedia Research Archive," this site illustrates the potential for the Web as a tool to present the work of a single artist, one who used many media to explore common themes throughout his work. Rossetti illustrated iconography such as "The Blessed Damozel" in print, drawing, poetry, and painting. Images and texts from many different collections and locations are brought together at this site within a critical framework that provides much improved access to rare and unique materials.

Intellectual Property

Using digital images in higher education has also been hampered by a lack of definition regarding intellectual property rights. By their nature, digital images may involve many layers of rights, pertaining to the original work, its photograph, a digital image scanned from that photograph, and (possibly) a subsequent digital image derived from that image. A common understanding of rights, permissions, and restrictions and a shared framework for administering rights would ease the burden of honoring intellectual property rights and further the educational use of digital materials. Such an understanding would have to be based upon broadly accepted terms and conditions for the use of materials.

Two initiatives are addressing intellectual property questions in relation to networked images. The Visual Image Archives Working Group of the Committee on Fair Use (CONFU) has been developing guidelines for the fair use of images in digital form that will aid those assessing the digitizing of existing image collections.[3] The Museum Educational Site Licensing Project (MESL; **http://www.gii.getty.edu/mesl)** brings representative U.S. museums, colleges, and universities together to explore the licensed use of museum images on campus networks. The partners in the MESL project are exploring administrative, technical, and legal requirements for the delivery of large quantities of high-quality museum images and information to educational institutions. Participants are also exploring the changes in the teaching process as a result of the incorporation of digital materials.

Opportunities

The MESL project has provided a vehicle for exploring and promoting the educational benefits of digital access to museum collections through campus networks. Faculty working on the project have been exploring the potential for expanding traditional ways of teaching and learning, and testing new ways to exploit access to both quality digital images and their accompanying documentation. Preliminary experiments seem to have proven successful in changing the classroom social dynamic as well as meeting curricular goals and teaching new skills.[4]

At the University of Maryland, coinstructors Sally Promey and Terri Gipps of the Department of Art worked with MESL materials to bridge a traditional cultural gap between the students in the art studio program and the art history program. Students from each of the programs were paired. The art history student was asked to write a description of one of the MESL images. This description was then passed on to the studio student, who created an image based upon it. The two students then compared the two works, and looked at the original, from the National Museum of American Art. Together both groups of students learned about the difficulty of communicating visual ideas in words, and built their awareness of issues involved in digital reproductive technologies.

At the University of Virginia, Professor Ben Ray used images of African masks from the Fowler Museum of Cultural History in a religious studies course. As a part of their term work, students were asked to create a virtual exhibition on the World Wide Web. This required humanities students, most often used to working and writing alone, to work in a laboratory setting and to collaborate. The resulting assignments showed a heightened awareness of the relationships between the aesthetic, cultural, and religious significance of these objects. The students themselves spoke enthusiastically about the experience, both in terms of learning and using new skills, and in terms of the changed learning process.

Conclusion

There are tremendous opportunities for the increased use of visual information provided through the Internet. New ways of teaching and learning can heighten the educational experiences of students. Networked communications can improve access to rare and unique materials and enable interdisciplinary exploration. What is required, however, for visual collections to be of utility to a new, broader audience is a critical mass of digital visual information that documents our cultural heritage. It is not enough that images be available in digital form; they must be accompanied by good description. If distributed resources are to be interoperable, they need to have been constructed within an overall framework that integrates both projects and collections through commonly held standards.

Notes

1. These issues, and others such as compression, are highlighted in Besser, Howard and Jennifer Trant, *Introduction to Imaging: Issues in Constructing an Image Database*. (Malibu CA: J. Paul Getty

Continued on page 21

The World Wide Web as a Resource for Scholars and Students

by Richard C. Rockwell
Executive Director, Inter-university Consortium for Political and Social Research (ICPSR), University of Michigan

In this wonderful annual meeting of social scientists and humanists, I assume it's not rude to talk about data, but it might be useful to define what I mean by "data." I refer to data of three kinds. One kind is the chart, table, or graph that you're accustomed to seeing. In the social sciences, these are called aggregate data. The second kind is administrative data, data taken from the records of government—raw data from which personal information has been stripped so that you can't identify a person even by guessing well. These raw data are available for analysis and often underlie those charts, graphs, and tables. The third kind is data taken from surveys, experiments, and other tools used by the social sciences to collect data. These raw data have been widely distributed among social scientists for several generations now and have formed the foundation for much research and education.

Access

The old technologies for disseminating those kinds of data were very labor-intensive, expensive, and slow. So when the Internet came along, there was a major motivation for the social sciences to begin to use it very quickly. Now we are beginning to see the original data collectors, the statistical agencies, and the data archives all taking a step beyond the dissemination of data—and that step is to provide access to the data. So rather than shipping the data sets across the network, you ship your query to the computer where the data reside, and then we give you an answer back over the network. If you want to see an example of how this works and what its potential is, look at the U.S. Bureau of the Census home page **(http://www.census.gov/)**. At the Census home page, you'll see the capacity to do a number of analyses on census data products, as well as to retrieve packaged tables and charts. This is the trend within all the statistical agencies of the U.S. government, some of them doing the work themselves, some of them "outsourcing."

In addition, as I mentioned, data collectors are beginning to provide direct access to their data. The Health and Retirement Survey and the Panel Study of Income Dynamics, each at the University of Michigan, are both offering very good, very user-friendly ways for people to get access to the data that they're producing: **http://www.umich.edu/~hrswww/** and **http://www.umich.edu/~psid/**.

At the Inter-university Consortium, we now have data from about 97% of our studies up for transmission over the network; that represents about 70% of the megabytes of the quantity of data in the archive. The archive—the largest social science archive in the world—totals a little upwards of 600 gigabytes of data. In addition we're offering new Web services, including search services that are tailored to social science data and social science data resources. We're offering analytical and extract services to the public on the National Archive of Criminal Justice Data home page, which you can find from the ICPSR home page **(http://www.icpsr.umich.edu/)**. We'll soon expand the scope of these public analytical services to data on aging and the aged. We're developing new archives that will be of interest to particular fields. One is going to be on education internationally, and we're working very hard to establish one on the Holocaust.

Documentation

In all of these things documentation is a major problem, for most documentation is still not electronic, including most documentation produced by federal agencies. What we get instead are what are called "data definition statements" for SAS and SPSS, which are two popular statistical packages. They don't document a study; they don't tell you how to use it, even for classroom use.

In order to put the contents of the ICPSR archive up for use, we have to scan around 350,000 pages of material—and I agree with Susan Hockey: OCR is not the way to do that. We're likely to adopt a proprietary product, Adobe Acrobat, as a standard for scanned documents; it does solve for us problems of large image files, and it partly gets around the OCR problem. But such documents are not logically structured: they are only images, or quasi-images. You need to be able to logically discriminate the name of the principal investigator from the code for missing data from the variables or universe statement. And none of that is possible with an image-scanned file.

What we need are structured documents, so that a study can be documented to a statistical package by the "code book" itself. Accordingly, we're developing something called a Document Type Definition (DTD) for SGML (Standard Generalized Markup Language) markup of social science code books, and we'll put that Document Type Definition in the public domain. Information on the DTD is available at the Data Documentation Initiative **(http://www.icpsr.umich.edu/DDI)**. We're striving to be compliant with the Text-Encoding Initiative (even though it wasn't really constructed for the social sciences). And we strongly

Continued on next page

Data

Continued from preceding page

hope to see federal statistical agencies use the Document Type Definition to prepare code books for their own data resources. If they do so, they will achieve a major improvement in documentation.

Work to be Done

Now, some observations on this brave new world into which we've entered: so much is enabled by this technology, it's sometimes easy to ignore what remains to be done. I'm going to take a perspective that could easily be identified as a "neo-Luddite" perspective. I think you'll see it's not quite that.

Recently, a faculty member at the University of Michigan said, "My students are now doing the research for their papers on the Internet and never go to the library!" That's a deeply worrisome trend, if it's widespread, because the Internet and the popular World Wide Web, the major tool on it, are really not yet mature information resources. The Web has some immensely valuable resources on it. But it also has a great deal that needs to be improved. We have very far to go before the words and the data currently available on the Web become information of the kind that we have grown accustomed to finding in our reference libraries and data archives.

Barely three years ago, the World Wide Web was the most recursive knowledge system that you could imagine: it was mostly about the World Wide Web! You'd click on a Web page and you'd get information about the World Wide Web. It had been developed at CERN in Geneva to solve a problem of internal communication, and it did that very well.

Even just three years ago, the Web was in its infancy, an amazing plaything for some enormously creative people. Now it's beginning to walk and to talk, and it's beginning to get into trouble, as toddlers often do. It has free speech issues, and there are performance issues. The development of "hot spots" in the commercial networks over the last few months is proving to be a major problem. Pricing issues will become a major topic of debate, with the old Internet ethic of "everything should be free" falling by the boards. Nothing is free; it's just a question of who pays for it. To reach its early adolescence successfully, the Internet must become technically stable. We can't have the situation in which home pages routinely disappear and in which "404 URL Not Found" is the message that you see at least a dozen times a day. It must be predictable or reliable so that whenever you go out looking for information from a single place you get the same information back every time—it's a little disconcerting when the information changes when it ought not to do so. That means that the Web has to have authoritative content. It means the Web has to be much more functional. It means that the Web has to have better response time. Further, it needs to be priced correctly, probably including pricing of the method of access. It has to be usable with a standard set of tools, including effective search tools, so that the user does not have to learn new procedures for every site that the user visits.

What are some means to these ends? Well, we need to have review procedures for content, not just for design. We need to have Web site reviews in the places where we now will also see book reviews. We need to ensure intellectual property protection and security for cash and information transactions. I think the disciplines ought to create explicit programs for identifying needs and gaps in the Web. There is a major program required to put the intellectual resources of our libraries online. The current digital library programs are just the barest beginning of what is required, although we have already made considerable progress in putting archival data resources online.

In achieving these ends, it seems to me that there is a professional group that we can use far more effectively than it has yet been used in the development of the Web: we need to put the design and implementation of the Web much more in the hands of information specialists, not in the hands of programmers. In other words, we need people who are able to do what librarians, cataloguers, archivists, editors, publishers, and book reviewers have always done. They know a lot, and their knowledge is germane to the development of the Web.

The anarchy of the Web is one of its more attractive features. Yet the resources are not available to permit endless duplication, reinventing of the wheel, pursuit of dead ends. It ought to be possible for major Web developers to work together to achieve common goals. The Coalition for Networked Information, the Web Consortium, and ACLS itself could offer institutional settings for such collaboration.

Our "Web Heritage"

Let me end by telling you what I think is a horror story: Web sites come and go every few seconds. The content on a given Web site can change one or two times within the course of a day. Thus it's somewhat frightening to realize that the Web is already a part of our national heritage, our cultural heritage. And as such, some aspect of its content must be preserved. Take the example of the statistical tables that will be available from the 2000 Census: The plans at the Bureau are to sharply reduce the printed publications program, preserving only the highly visible works, such as the

Continued on page 21

The American Arts and Letters Network (AALN)

by Charles Henry
Vice Provost and University Librarian, Rice University

The American Arts and Letters Network (AALN; **http://www.aaln.org**) is a Web-based project sponsored by the ACLS, the Coalition for Networked Information, the National Initiative for a Networked Cultural Heritage, and funded by the Mellon Foundation. It is an experimental Web site designed to enhance teaching and scholarship as well as to preserve and make more accessible digital examples of the global cultural heritage. The AALN collects networked resources in broadly defined areas of the humanities, arts, and some of the social sciences for a wide range of constituencies, including young people.

One distinguishing feature of AALN is that all data sources are reviewed for quality of content by an advisory board before becoming linked to the network, in an effort to construct a reliable and vital resource.

Why the AALN?

The American Arts and Letters Network was established in response to some of the many problems associated with the World Wide Web. These difficulties include its size (estimated at nearly 30 million sites as of this writing); a paucity of sophisticated protocols (HTML is very simple, which is one reason it has proliferated); also, millions of sites on the Web are constructed in isolation, which hinders standards and leads to redundancy of effort; a heavy reliance on images, which often tease the eye more than delivering substantive information; an increasing number of commercial sites; and many out of date sites that are no longer maintained but might appear at first glance to be current.

Other concerns include the firewalling of information to those outside of an organization or institution, rendering some excellent sites inaccessible to most of the user community; misleading titles (no one can really tell the substance or value of content of a site by the title, which is the only way at present to get to a Web page); there is little academic peer review; and, perhaps more philosophically, the World Wide Web so far tends to be used anonymously and passively.

Structure

The structure of AALN is straightforward: because graphics capability is by no means universal, the pages are purposefully without pictures, moving or still, and relying on easily read text to convey the message. The sections of AALN include information on scholarly societies, linking many organizations relating to education; information about the World Wide Web and how it can be effectively used; and "collections": the largest section devoted to Web sites deemed of value in a variety of academic disciplines. For ease of use, the project has adopted the disciplinary focus of the ACLS societies.

The program is easy to use, self-explanatory for the most part, with the option of simple or advanced searching of the linked sites. As the project develops, forms will be made available for those desiring their pages to be linked to AALN, with the criteria for inclusion stated on the electronic document. Evaluation forms will also be introduced in the coming months in order for Web site creators to learn of users' response to their work.

Goals

In light of the current configuration of the Web, and the behavioral patterns associated with its usage, the two main purposes of AALN can be summarized as follows.

The first is to provide a reliable place to go for looking up current, well maintained information pertinent to the humanities, arts, and social sciences. This project, it should be stressed, does not attempt to create new information on the Web, but rather seeks and then links those sites deemed applicable to its user community. Many excellent sites exist, such as Voice of the Shuttle, Georgetown's medieval database "Labyrinth," Internet Music Resources, sites for electronic journals, and key electronic text centers such as those at the University of Virginia and the Center for Electronic Texts in the Humanities (CETH) at Rutgers/Princeton.

AALN in no way attempts to duplicate this fine work, but rather knits them together in an easily accessible web, a starting point that avoids the often enormous hit rate and mixed-value content of the larger search engines such as Lycos and Alta Vista, which can generate tens of thousands of hits on a single term, most of which are not applicable to the researcher.

The second goal is perhaps the more important, and lies in the future. The inherent facility of the Internet and the Web for collaboration is at present underutilized. The designers of the American Arts and Letters Network hope this project can eventually foster a collaborative presence on the Web, a "place" where teachers and scholars and students can go to discuss issues, exchange ideas, participate in online conferencing, and collaborate on projects of mutual interest. This activity would be supported by archived hyper-mail, MUDs and MOOs, and other spaces that serve as a forum for the exchange and recording of ideas.

Another facet of this goal is to establish in coming years a genuine peer review and ongoing evaluation process of networked information that is pertinent to the humanities: to better understand how these networked resources can be used in teaching and scholarship, how they might transform current methodologies, and to assist in establishing workable standards for future online resources.

Continued on page 22

The National Initiative for a Networked Cultural Heritage (NINCH)

by David Green
Executive Director, NINCH

Introduction: Technology, Scholarship and Politics

The National Initiative for a Networked Cultural Heritage (NINCH; **http://www-ninch.cni.org**) is a membership coalition established to assure the fullest possible participation of the cultural sector in the new digitally networked environment. ACLS was a leadership sponsor in founding NINCH, recognizing the urgent need for such a broad-based organization to meet the challenges of the new technology for scholarly communications.

The organization by ACLS of the panel "Internet-Accessible Scholarly Resources for the Humanities and Social Sciences" for its 1996 Annual Meeting shows its continuing leadership in examining how useful (indeed, how transforming) the new technology can prove to the work of humanities scholars. ACLS leadership has in fact done much to position this community so that it can begin to consider how to assess and influence the development of the technology and the new social, political and economic arrangements it will bring in its wake. These will influence not only scholarly communications but the very way our society learns about itself. Beyond this, ACLS has sought to build, through NINCH, a broader coalition that would facilitate joint research and action, beyond the scholarly community, on both practical and political levels to ensure that we reap the most benefits from the telecommunications revolution.

One critical meeting in this process was the 1992 Irvine conference, "Technology, Scholarship and the Humanities,"[1] organized by the Getty Information Institute in collaboration with ACLS, the Coalition for Networked Information, the Research Libraries Group and the Council on Library Resources. There, leaders across the field began to collectively consider the scholarly implications of the digital revolution. A year later, when the Clinton Administration's 1993 Action Agenda[2] for creating a National Information Infrastructure failed to acknowledge any role in the envisioned NII for the arts and humanities, a leadership group was assembled in Washington, DC, to organize the community's broad response. NINCH is one result of that effort.

The National Initiative was thus formed at the nexus of scholarship and politics. Linking the activities and concerns of the scholarly community with those of the larger cultural community can only strengthen both as we work to educate our own field and to influence national policy. Museums, libraries, scholarly societies, research and educational institutions and contemporary arts organizations are joining NINCH to help build an environment in which people and institutions can be encouraged and supported in networking their cultural resources in smart and integrated ways.

The State of the Field: Vision and Needs

Much good and constructive work is being done in many fields in what we might call the first phase of the massive project of making cultural heritage materials widely accessible in electronic forms. Resources are being built, whether they be hyperlinked scholarly and historical editions (e.g. the Model Editions Partnership projects: **http://mep.cla.sc.edu**,[3] the Brown University Women Writers' Project: **http://www.stg.brown.edu/projects/wwp/wwp_home.html**, or Multimedia Yeats at the University of Wisconsin: **http://www.oit.itd.umich.edu/WIP/ArtsHumanities/MultimediaYeats.html**), virtual research archives on a single figure or topic (such as Jerome McGann's Dante Gabriel Rossetti site: **http://jefferson.village.virginia.edu/rossetti/rossetti.html**), or guides and gateways to Web-based resources (such as Georgetown's "Labyrinth" guide to medieval studies: **http://www.georgetown.edu/labyrinth**, or the collaborative Vergil project at Pennsylvania: **http://ccat.sas.upenn.edu/~joel/courses/project.html**), experimental teaching tools (like Steven Marx' Blake Multimedia Project: **http://www.fmdc.calpoly.edu/libarts/smarx/Blake/blakeproject.html**) or collaborative multidisciplinary research tools (such as John Dobbins' Virginia-based Pompeii Forum Project: **http://jefferson.village.virginia.edu/pompeii/page-1.html**).

But even the broadest of these, those that sort through, collect and direct Web-navigators to collections of materials, are single projects. They still have walls, and boundaries, idiosyncrasies and resources that freeze or disappear without trace. And the Web, despite its multimedia capability and the power of hyperlinking, is still a rudimentary tool for deep scholarly research; many electronic projects cannot be networked.

Our vision of what is possible through networked cultural heritage is of a unified catalogue, a universal library, an integrated, distributed body of cultural material, in which both simple queries and complex research questions can pull richly documented resources from a plenitude of locations and display them on the desktop. Such a library would be widely accessible, seamlessly interoperable across many media, of the highest possible quality and fidelity, and easily usable and searchable—by creators, scholars, the general public, and by teachers and learners of all ages.

The task before us is both technical, social (development and acceptance of a range of technical and procedural standards), economic, and institutional (both in costs, cost-structures, and in

NINCH

Continued from preceding page

the reorganization of everything from business practices, to institutional structures and the way we teach and learn). It will need to engage not just technicians, engineers, and software designers but, more importantly, scholars themselves, teachers, librarians, archivists, "information professionals," curators, artists, administrators and good thinkers, whatever their particular hat.

Libraries and archives are doing a good job in beginning the necessary work of preparing the ground for coordination and integration of activities, projects and standards. For example, the National Digital Library Federation **(http://lcweb.loc.gov/loc/ndlf)** is providing an infrastructure for libraries to cooperate within. The National Digital Library Program **(http://lcweb2.loc.gov/ammem/ammemhome.html)** at the Library of Congress is deliberately tackling the hardest issues it can find in digitizing material so the challenges can be clearly defined for the rest of us. The American Heritage Virtual Archive Project **(http://sunsite.berkeley.edu/FindingAids/EAD/ameriher.html)** is an experimental prototype to make available in a single searchable database hundreds of finding aids to primary source materials in American history and culture from collections at four major research libraries.

All of these are good beginnings. The charge and imperative for NINCH is to encourage and catalyze further coordination and integration of effort. We must extend and share the activities and lessons learned by universities, libraries, museums and others in the context of a complex political landscape, which the current copyright debates brilliantly illuminate.

The Start-Up Strategy

NINCH is now in the midst of a start-up strategic plan that reflects three paramount concerns:

- that the cultural community have a clear sense of direction about its role in the new digital environment so that constituencies can develop a coordinated, systematic approach to networking cultural resources;
- that government and industry understand the place and potential of the arts and humanities in the development of a Global Information Infrastructure;
- that this very broad and diverse community have a common, dependable communications infrastructure it can use to share and learn about developments in digital networking as well as to coordinate its positions and expertise on issues, research and developments in the field.

NINCH's immediate solutions to these concerns are, respectively:

Community Building: We will include as many members of the cultural community as possible in the task of building a platform where we articulate and voice a clear direction in the digital arena.

Advocacy: NINCH will make the case both to government, to other major institutions and to our own constituency about the critical importance of ensuring that the arts and humanities community be a partner in the construction of the National Information Infrastructure. As a start, it has created a working group to plan a 1997 Congressional campaign and to develop our long-term advocacy strategies.

Communications and Education: NINCH is constructing its own communications infrastructure on which it will gather and distribute news and information about networking developments, provide guidance and leadership in prioritizing the issues we face and communicate a broad working agenda to our member organizations.

Our listservs and Web site **(http://www-ninch.cni.org)** for announcements, newsletters, and resources are operating, and we invite your participation. We have established working groups in advocacy, copyright and fair use policy, and a cultural diversity project, and two cross-sectoral research projects are under way.

Innovation is a social and cultural phenomenon, not a technological one. With much technical work to be done, the cultural community needs a strong and broad enough vision of itself and its needs in this area to be able to articulate and command what it wants, rather than riding sidecar to the main show. As a coalition, NINCH has immense potential to catalyze enough synergetic partnerships to carry us forward in the right direction. We trust the scholarly community will continue to be a full partner in this endeavor.

NINCH Members

American Association of Museums
American Council of Learned Societies
American Historical Association
Americans for the Arts
Art Libraries Society of North America
ArtsEdge/John F. Kennedy Center for the Performing Arts
Arts Wire/New York Foundation for the Arts
Association of Art Museum Directors
Association of Research Libraries
Coalition for Networked Information
College Art Association
Commission on Preservation and Access
Council on Library Resources
Getty Information Institute
Library of Congress
Museum Computer Network
National Association of Artists Organizations
National Assembly of State Arts Agencies
Research Libraries Group
Society of American Archivists
Smithsonian Institution
Visual Resources Association

Continued on page 22

Because It's Time: A Commentary on the Program Session

by Willard McCarty
Editor, "Humanist"
Centre for Computing in the Humanities, King's College

Two broad themes arise from the presentations in this session: first, that the time has come for critical assessment of the Internet as a medium for scholarly communications; second, that (in the language of programming) we are beginning to glimpse the difference between the *bugs* and *features* in our uses of this medium. It is time because we now have a sufficient mass of scholarly resources and enough direct involvement with the Internet to attempt a sorting of characteristics from the confused mass of temporary infelicities and misuses. To put the matter another way, we should now be able to throw off the imitative thinking typical of the early stages of a new technology and try to understand where the real genius of the medium might lie.

The panelists—Susan Hockey, Jennifer Trant, Richard Rockwell, and Charles Henry—spoke with notable unanimity about the limitations of current tools and materials. Nevertheless, as Trant pointed out for art historical images and Rockwell for social science data, the bare fact of their accessibility is of enormous importance, however poorly we have supplied and used them to date. (A comparison with the early years of printing would be instructive here.) If mere accessibility were not so important, there would be much less reason for us to turn our attention prominently to them and detail their faults by our most rigorous standards. Thus, for example, while we are cautioned to beware of search results from badly transcribed texts and to regard the World Wide Web as a mere beginning, both are useful within their limits. More importantly, they show that new tools lead to new questions, and begin to suggest what those new questions might be. Dismissive criticism is as senseless as uncritical acceptance.

Another perhaps obvious but vital point concerns our fitness for the task. All too often humanists and social scientists of the less numerically intensive sort imply in their remarks about technology that we are incompetent or at least inexperienced. Thus either we are portrayed as poor suppliants at the industrial/scientific banquet, or the introduction of computing into humanistic scholarship is viewed as the incursion of an essentially foreign, possibly inimical force. The truth is, however, that poets and philosophers have been thinking and dreaming about automata (the general class to which computers belong) at least since Homer.[1] It is only within the last few years that computers have become sophisticated enough even to begin to handle our data and so be worthy of our attention. As Northrop Frye pointed out in one of his last public lectures,[2] humanists have in the past distinguished themselves by their powerful applications of "advanced" technology, such as the codex or movable type. So, we need not worry about our qualifications.

Remarks of the Panelists

In her remarks, Susan Hockey wisely defined the central function of the Internet as remote access to material, not as whatever the World Wide Web and its HTML may currently allow. If we look critically at the world defined by the Web, or indeed as defined by current offline software, we can see its radical insufficiency to meet the needs of basic research in the humanities and social sciences. Thus, Hockey notes, the profound complexity of humanities data and how poorly it is served by the primitive retrieval methods available on- or offline. In other words, we cannot rest content, nor be seen as such by our more skeptical colleagues, but must rise to the challenge that these data offer to systematic treatment. Here lies a great opportunity for scholars to help solve problems which extend well beyond their specialist areas and conventional methodologies.

In a time when superstitious reverence for learning no longer shields us from public skepticism, it is vital not to overlook any opportunity for demonstrating the ultimate usefulness of pure research. This usefulness may at least be argued from the general principle that to solve a problem in a more complex form is to solve it in many simpler manifestations. If, that is, scholars in the humanities and social sciences bring knowledge of their complex data to the task of developing the software tools, demanding that *these* data be the yardstick, then we are far more likely to see some advancement of the medium than if processing the language of weather reports or medical diagnoses is the goal.

In a related point, Hockey has emphasized the crucial question of *how these data are used* in the usual course of scholarship. One rather new aspect that software brings to humanities and social science data is that its manner of use becomes inseparable from providing access. This has a number of important consequences. Socially, as Jaroslav Pelikan points out in *The Idea of the University*, the change means that it is becoming unworkable, and even more mutually damaging than formerly, to segregate those who support research into a lower, servile class; in the common interest of our survival, they must be made colleagues in the research enterprise. Intellectually, among many other things, it means that the central task of software development for the humanities and social sciences, whether on- or offline, is the modeling of scholarly methods. This modeling has been and

Because It's Time

Continued from preceding page

continues to be done by an often productive alliance between scholar and programmer. I would argue, however, that the ideal lies precisely where the evolution of computing appears to be taking us: to the point at which the scholar constructs his or her own tool, directly, by assembling methodological "primitives" into a process that models thought. One might call this the Lego-approach, or to use a term coined by American physicists from my own generation, the "tinkertoy modeling" of scholarship. There are two points to this sort of modeling: first, to produce results; second, and primarily, to learn more about our data from the failures of the model.[3]

Hockey notes that at least for now textual scholarship seems to be better served by the techniques of markup than by clever algorithms. Her perspective is typical of the computing humanist, less so of the computer scientist or computational linguist, whose algorithms we continue to hear about but seldom if ever get to use. Markup, though laborious to encode, has the advantage of putting into the hands of the average working scholar the means of rendering human intelligence into a form that even simplistic software can process. No large grants are then required to continue the work, and the intelligence of the scholar is engaged with the data in a direct, immediate way. Literary scholars will recognize this as a new form of "close reading," and in it philologists find their ancient passion.

Textual markup is often misunderstood, and its potential grossly underestimated, by the tendency to think of it as belonging to a preparatory, mechanical stage one must get through before serious work can be done. In some cases, of course, tagging is largely mechanical, but the more complex the phenomena one wishes to identify consistently, and so be able to process reliably, the stronger its role as a means for manifesting one's understanding of the data. (Give the job to a graduate student and you have just made him or her a colleague.) Tagging, I like to argue, is akin to translation and so forces deep knowledge as much or more through its failure to convey meaning as by its successes. Tagging is the form of tinkertoy modeling most readily available to those with the most intimate knowledge of humanities and social science data. It is an instrument of perception, a medium for thinking about text.

Markup also provides for the encoding of metadata, the editorial information about information that allows the scholar to document the history of the text, its status, and what he or she has done to it. Without this meta-information the electronic book is, as it were, without a title page and preface.

The potential of markup is crudely and poorly represented by HTML, which is a highly simplified form of the Standard Generalized Markup Language (SGML, now adapted for scholarly purposes by the Text Encoding Initiative, thus TEI/SGML; **http://www.uic.edu/orgs/tei/**). Currently the preferred way of delivering primary and important secondary texts across the Internet is to render SGML-encoded text on demand into HTML, using so-called "gateway" software. This has the advantage of allowing us to exploit current Web technology without requiring that we make a large investment in tagging our data with a still primitive and severely limited metalanguage. As HTML improves, or when it is discarded for something better, only the gateway software need change, not the tagging of the data.

Scholarly publications solely in electronic form are becoming increasingly visible. Again, the imitative mode of thinking still seems to prevail in many of these, but gradually we can observe signs that the unique or at least characteristic potential of the medium is being realized and exploited. Hypertextual, or more generally, hypermedia links to other resources are the first sign of something genuinely different, especially when these link within a document to remote software. When the reader of an online article can, for example, access the database on which the research is based, directly from the article, then we glimpse something new—or not so new? Is this not, interestingly, a mechanical model of an educated reader's memory, one further step in the direction begun by the external, written record?

Until very recently, computing in the humanities and social sciences has on the whole paid much less attention to graphical data than to verbal or numerical, for the simple reason that images require better hardware even for the simplest operations.[4] The World Wide Web has, of course, thrust image-delivery into the forefront, but as Jennifer Trant pointed out, such delivery is still in a primitive state. There are, first of all, a host of technical problems, e.g., with the control of color: the provider has in fact no precise control over the actual colors that the viewer will see. Image files tend to be quite large, even with current compression techniques, and so present a serious problem for timely delivery. More difficult is finding images in the first place. The best we can currently do is to attach keywords to each image-file in a database, and provide small versions of the images (called "thumbnails") on a Web page for browsing. We are still, she notes, going to particular and often idiosyncratic collections, not to the images directly through some kind of central

Continued on next page

Because It's Time

Continued from preceding page

index or finding tool. We are still very far from usable image-recognition software, which would allow scholars to search by shape.

The print medium has had other difficulties with images that the online medium avoids, however. Printing them is expensive, integrating them with text crude and often difficult, and the final arrangement is necessarily fixed. If we can assume that the user already has adequate equipment for viewing color images—taking due care to compensate for "developed-nation myopia"—then cost is no longer a problem, so that they become practical where they were not before. Furthermore, images can follow the text, or the text follow the images, as the reader prefers, and detailed commentary (verbal or pictorial) can be attached directly to the visual details. Art historians are, of course, in the middle of the action, but scholarship in many other areas is bound to be affected if for no other reason than the ease of including images in text. Furthermore, even the art historians have, as a rule, not concerned themselves much with the techniques of imaging, whereas now they must at least be acquainted with these techniques if not be capable of applying them skillfully. The disciplines mingle, as for example in the Piero Project at Princeton University **(http://mondrian.princeton.edu/art430/art430.html)**.

Scholars in the humanities and social sciences have a great deal to learn about imaging—not only the technical aspects, but even more importantly for those outside art history, the relevance of images to scholarship. Some literary scholars, for example, tend to regard them as unnecessary decoration or entertainment, a distraction rather than an aid to understanding. My own opinion is that we do not know very clearly when they are distraction, when essential. The online medium urges us to find out.

In *some* respects, social scientists are ahead of humanists in their applications of computing. Perhaps simply by virtue of longer and more widely accepted involvement, social scientists have, for example, been quicker to recognize the importance of the online "data center," where one may get easy access to large amounts of information, e.g., of demographic or economic data. In some universities, for example, undergraduate students can easily share use of primary data with their professors and perform sophisticated analyses as part of routine classwork. Ideally, and perhaps also in practice, such access is or will result in a much more informed populace, one much less easy to fool by misleading statistics drawn from restricted datasets. This at least is the chief social aim, and it clearly illuminates the importance of access.

Rockwell spoke to these points, but he also returned to a point made by Hockey, that the "information specialist" plays a key role in making this access meaningful. Librarians of the dawning present and immediate future are increasingly going to need scholarly knowledge of how the data may be used and abused. Their training needs to be methodologically informed. Their task would be more than daunting, perhaps even impossible, were it not for the fact that from a methodological perspective, much of what scholars in the humanities and social sciences do belongs to a common ground. Inevitably, it seems to me, as the library merges with the computing center, it also becomes something much more like an academic department. The nature of the closer collaboration between computerized libraries and academic departments must still be worked out, but I think we can hazard a guess that the research-methods course is the natural point of contact, as it has sometimes been in the past. Here is an excellent opportunity for "collegial service."

Charles Henry focused on two primary issues that nicely complete the picture I am drawing here. His topic was the American Arts and Letters Network (AALN), which he defined as an experimental, refereed Web-presence aimed at assessment of online resources in the humanities. Allow me to conclude my review of the panelists' remarks by deconstructing his definition as a series of comments on the subject of the meeting. I focus on three terms: *experimental*, *refereed*, and *assessment*.

Experimental. This implies, of course, something tentative, unfinished, to be valued more for what it discovers in the attempt than for its material accomplishments. An experiment puts an idea or hypothesis to the test, and thus looks to the future. Computer technology, a friend has observed, is always in the future tense; what we have is never as good or important as what we are about to get. The conclusion I draw from my attempts to teach humanities computing is that particularly in this incunabular phase, the ideas are what really matter because they are what survive, perhaps even become clearer, from one version of a computing system to the next. So, we learn from an experiment, then move on. As with questions, a good experiment leads to a better one, not to a final answer that, as Blake saw, puts the light of knowledge out.

Refereed. As any careful survey of online scholarly activity will show, refereeing is now commonplace, although it should not be assumed that non-refereed items are necessarily less serious. The experimental nature of the medium implies that many are serious if for no other reason than they show us something new. Further-

Because It's Time

Continued from preceding page

more, in some cases online publications have other means for guaranteeing the quality of what they publish, e.g. in highly specialized, sparsely populated fields where assessment is implicit. Nevertheless, it should be abundantly clear that online publication is now sufficiently popular among scholars that filtering mechanisms are needed. Cost is in general no longer a factor, but individuals must have an intelligent way of deciding quickly whether a publication is worth reading. Editors of online publications should be highly motivated to set up good refereeing mechanisms, since they have to convince both conservative readers and often even more conservative authors that they are worth the attention.

Assessment. Referees assess, but here I refer to the prior meaning of assessment in the incunabular phase of online publication: discovering what the standards should be. It is not to be assumed that these are simply identical to those we have used in the past. They may be quite different.

First we need to discover the particular genius of the online medium. What is it good for, what is it not good for? Journal editors and publishers most often see the urgent matter as economic, the electronic medium offering survival in the face of certain extinction. A deeper question, however, is the uncertain balance between what the medium can be made to do and what it is particularly suited for. Where progress intervenes, to solve a problem that formerly looked like an inevitable characteristic of the medium, is not always possible to say. If, for example, bibliographic instability and other forms of transitoriness are basic to the medium, then it is foolishness to pour effort into attempts at permanence. Can we rely on progress to solve the problems of instability and transitoriness? To me the answer is not obvious.

I hasten to stress that even in such a highly technologized area, few of the issues are purely technical. Even if some do not require our particular skills, they impinge on scholarship in significant ways.

Once the characteristics of the medium are well separated from the temporary problems, it will be much easier to see how it might best be used. This is not simple, even if the characteristics are known, because scholarly publication is an intimately interrelated part of our disciplines. *The problems we face are systemic, not isolated.* If, for example, relative transitoriness turns out to be intrinsic to the medium, then online publication is best suited to the quick exchange of ideas now more characteristic of the social sciences than the humanities. Were the humanities to take on the new style, then the nature of the work might well change, and with it the ways in which professional assessment is done, the criteria by which people are recruited, and so the kinds of individuals in our departments and eventually the nature of the discipline they practice. As I pointed out earlier, online publications can offer direct access to the data on which the research was done, e.g., a tagged literary text. To the degree the opportunity is exploited in literary critical writing, the nature of this writing will change, and with it the discipline, and so on.

An analysis such as the above runs the danger of reducing a highly complex, contingent, and so ultimately unpredictable situation into a set of known forces whose interactions we think we can control. It is salutary to consider, as Robert Frilled has in his recent book *Zipper: An Exploration in Novelty*,[6] just how contingent new technologies can be. With some, such as the telephone, individuals have from the beginning been able to foresee their application and significance, but with others this has emerged much more slowly, or as with the zipper, their very survival has remained uncertain for a long time. We must therefore be cautious in adopting visionary pronouncements—the new medium is especially prone to these—but at the same time quick to get involved. As has been pointed out, the new medium is rapidly being shaped and is rapidly shaping the world around us. Our intellectual condition if not survival appears to depend on our engagement with this exciting adventure in scholarly communication.

A Provisional Typology

Clearly much is happening. How are we to fit it all together and make it work effectively for scholarly purposes?

Here I would like to sketch in broad outline how I think our engagement in the construction of a culturally rich Internet might work. My suggestions, which I intend only as provocation to debate, take the form of brief commentary on the diagram that follows, "A Typology of Activities for Scholarly Inter-net Resources" (page 18). I invite those with a more intimate knowledge of the political and social realities, or simply with better ideas, immediately to begin modifying my sketch, or to replace it altogether.

The fundamental problem, I think, lies in focusing and coordinating activities into a coherent effort. We have no lack of talent. In some countries, like the U.K. or Norway, centralization of the universities and of research funding makes this a considerably simpler task than in the U.S., where (at least to this outsider) a bewildering array of organizations and institutions appear partly to compete, partly to cooperate. Countries other than the U.S. are particularly important to keep in mind, not only because they are different, but chiefly because the Internet is by nature international. The fact that English is currently the *lingua franca*,

Continued on next page

Because It's Time

Continued from preceding page

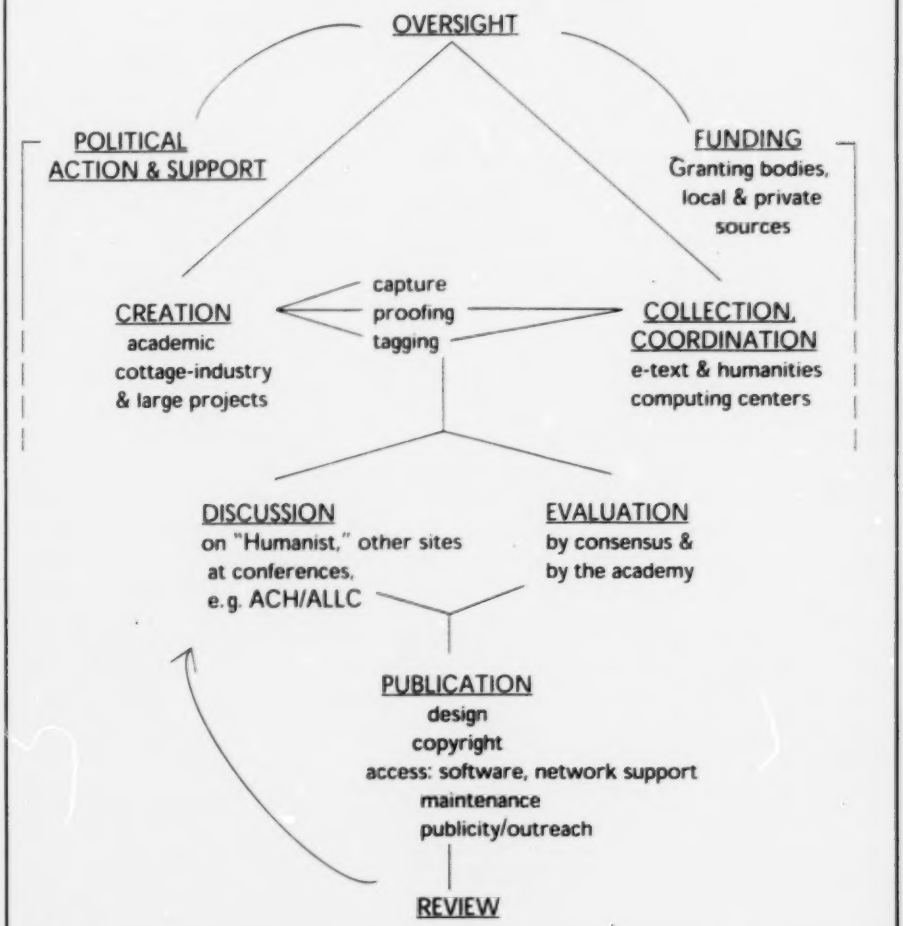
and that the U.S. is sufficiently powerful to maintain the illusion of self-sufficiency, is a genuine impediment to a full realization of what the Internet can do for scholarship.

Now to the diagram. I begin with the core-function of academic governance or **oversight**. Especially in the U.S., with its number and diversity of players, leadership and coordination are crucially necessary to bring together all the players and persuasively to suggest how they might cooperate, what the best agenda might be. In essence the academy as a whole, however it is constituted, needs to take charge. In the U.S., the ACLS is very well positioned for the role; in the U.K., for example, the British Academy, perhaps also the Arts and Humanities Data Service **(http://www.kcl.ac.uk/projects/ahds/home.html)**, seem the logical counterpart.

To one side of this central role, **political action and support** are continually necessary to foster and maintain public approval for our initiative to transform the Internet into a cultural, or better, intercultural resource. This approval is obviously vital for adequate **funding**, through national agencies such as the NEH in the U.S. and SSHRC in Canada, as well as through private and local sources. Fundamentally, I would argue, what we need can only come about if we cultivate an informed populace. At least in Canada and the U.S., it no longer seems possible, even if it were desirable, to maintain an academy in monastic isolation from the world. Increasingly, protection from the heat and dust of ordinary life is a privilege constantly to be justified.

Direct communication between academics and the tax-paying public is difficult, however. Ursula Franklin, a renowned scientist at Toronto, makes the point in a recent remark that, "As students we were warned: Only the great dare touch the commonplace."

A Typology of Activities for Scholarly Internet Resources



Most academics lack the rhetorical skills needed for successful communication with the public—chiefly the ability to speak in effective oppositional sound-bites—and are not likely to gain them. Thus political action and support require specialists, for example in the U.S. the National Humanities Alliance. I hasten to point out, however, that our students already provide an essential channel of communication to the immediate future; they constitute especially fertile soil for ideas about the application of computer-mediated communication to humane ends. (Thus although good teaching is not on my diagram, it is everywhere implicit.)

So much for support. The next matter is the acquisition of material for a scholarly Internet. Partly this involves **creation** of new electronic data, partly the **collection** of existing data and **coordination** or organization of the results. Creation is laborious but best suited, I think, to the academic "cottage-industry." This industry consists of funded projects, large and small, but also of independent efforts by academics who may produce resources as a by-product or co-product of their research. Realizing the potential to any significant degree requires, however, that researchers understand that this potential exists. They may not, because it is largely a new phenomenon.

Because It's Time

Continued from preceding page

By nature computer-assisted research puts into readily communicable form supporting materials that formerly were lost or at best preserved as "foul papers." If our colleagues can be persuaded where appropriate to approach the construction of these materials systematically, with a view to their use by others—to think of them as adjunct publications—then much will be made available for the Internet without additional cost. Key to the process are the e-text and humanities computing centers, because they provide the advice, training, and means of preservation. The Summer Seminar, held annually at Princeton University by Hockey's Center for Electronic Texts in the Humanities (1991-96), performed an essential service in this regard by educating the educators. Also necessary are suitable venues for these adjunct publications."

Once resources have been produced, especially in the incunabular stage we currently enjoy, they must be viewed as even less finished than a conventional manuscript submitted for publication. The standards by which to judge such a manuscript are well understood; those applicable to electronic resources are not, as I have suggested. Thus we need intensive, critical, intelligent **discussion**. This is already taking place, for example on "Humanist" (**http://www.princeton.edu/~mccarty/humanist/**), in more specialized online seminars, and at conferences, such as the annual joint meeting of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing (ACH/ALLC; **http://www.ach.org/**). Here all we need do is to recognize the activity and more officially to engage in it by bringing in the scholarly organizations, and through them their members. Insofar as possible, the entire community must participate.

Out of discussion comes the means for **evaluation**, informally through a consensus that arises from discussion, formally by the academy at several levels. Within institutions it can only begin once the value of working with electronic materials is properly recognized. This may be quite difficult without persuasive exemplars, general agreement on standards of work and means for evaluating it. Again, the academy itself, through such organizations as the ACLS, is the logical place to look for help in putting together what we know and assembling examples of good work. In recent years, at several academic gatherings, the need for a directory to exemplary computer-assisted research projects has been strongly noted. We have the means of communication successfully to construct such a "bibliography" and make it available; the (admittedly large) task only needs to be done.

Once evaluated, work proceeds to **publication**. As one publisher recently noted, "We all agree that electronic publication eliminates the middleman, but we disagree as to which one of us is the middleman." Among academics, university presses are not infrequently fingered, but it seems naïve to assume that we can simply take on the impressive set of skills a good publisher has to offer: among others, editing, design, legal negotiation, publicity—and validation through the *imprimatur* of the press, signifying quality of contents. Furthermore, the middleman argument implicitly assumes that electronic will replace print publications, i.e. that the former are the same kind of thing as the latter, only cheaper. As I have argued throughout, this is to miss the essence of the matter entirely.

One cannot help noticing, however, that the private "electronic press" is now very easy to establish and operate, and that a prodigious flowering of new electronic publications is daily stretching the limits of conventional forms. These feed back into the ongoing discussion, where we must ask steadily, "What works? What does not work?"

Finally in my diagram is the **review** by which published products are judged, and so publishers and their authors kept honest. There is little to say about this, except that the need for informed reviewers also drives the evolution of the system. It is still true that in this time of flux and few jobs, we must reach for talent wherever it can be found, and in that reaching lies opportunity for the highly trained but under- or simply unemployed.

Conclusion

Where from here? Such meetings as this one give considerable hope that we are on the right track. Again, I think it essential that we understand our fitness for the task, how much our skills are needed, and the crucial role our intelligent participation will play in our own long-term survival.

As serious books give meaning to the older technology of printing and binding, so high-quality, culturally rich Internet resources will provide leadership by showing what the newer medium can do. Thus the social, outward-looking aspect of the job before us. The intellectual, inward-looking aspect is, I suppose, obvious to everyone by now. Particularly with the advent of the World Wide Web, crude though it is, the utility of the Internet is clear for all kinds of scholarly work.

Egoque ipse multa quae nesciebam scribendo me didicisse confitear, Augustine wrote, "and I would confess that there are many things I did not know that I have learned in the course of writing" (*De trinitate* 3.1).[8] Writing, he saw, is a means of perception and instrument of thought. I made this claim earlier about textual markup, but now I wish to extend it to computer-mediated communications as a whole, and so to assert in closing that the fundamental subject here is the transmigration of thought, or less philosophically, the refurbishing of our cultural heritage. It's in our job description.

Continued on next page

Because It's Time

Continued from preceding page

Notes

1. Homer, *Il.* 18.376f; *Od.* 7.91-4. See also Apollod. 1.9.26; Apoll. Rhod. 4.1638-93 (cf. Frazer's edn. 118-19n.). I discuss automata in "Language, Learning, and the Computer: desultory postprandial investigations", in Peter Liddell, ed., *CALL Theory and Application* (Victoria, B.C.: Univ. of Victoria, 1993): 37-55.

2 "Literary and Mechanical Models", in Ian Lancashire, ed. *Research in Humanities Computing 1*, Papers from the 1989 ACH-ALLC Conference. Oxford: Oxford University Press, 1991.

3 See my discussion of modeling in "Encoding Persons and Places in the *Metamorphoses* of Ovid: 1. Engineering the Text", Text, métatexte, métalangage, ed. Brian T. Fitch, *Texte: Revue de Critique et de Théorie Littéraire* 15/16 (1994): 278ff. For information about *Texte*, see **http://www.chass.utoronto.ca:8080/french/litera/Revue_Texte/Texte.html**.

4. See the excellent *Introduction to Imaging: Issues in Constructing an Image Database*, by Howard Besser and Jennifer Trant, (Malibu CA: J. Paul Getty Trust, 1995), also available at **http://www.gii.getty.edu/intro_imaging/home.html**.

5. See, for example, the Data Centre, Computing in the Humanities and Social Sciences, University of Toronto, at **http://datacentre.chass.utoronto.ca/**.

6. New York: W. W. Norton, 1994.

7. With Russon Wooldridge I edit one such venue, particularly for the computer-assisted methodological aspects of research, *CH Working Papers*, published online at **http://www.chass.utoronto.ca/epc/chwp/** and at **http://www.kcl.ac.uk/kis/schools/hums/ruhc/chwp/**.

8. I am grateful to Professor James J. O'Donnell (Classics, Pennsylvania) for putting these words before me. □

Building the Scene

Continued from page 2

as images and then combined with text (as in the JSTOR and other electronic journal projects, where keyword searching of journal articles is done on "good-enough," but not 100% accurate, OCR scanned text and the found text is presented as an image of the actual journal page). Either way the scanning procedure is fairly simple: the major challenges lie in the encoding of text and in developing software for analysis of the resultant e-text.

Images can only be digitized through one means: a scanning camera. However, cameras differ from one another in quality, scanning resolutions vary tremendously and, once digitized, images are currently compressed in different ways for transmission and will be displayed on monitors with different characteristics. Establishing quality control is critical.

More has to be done both in gaining broad acceptance for consistent technical standards as well as for standards in describing and cataloging resources. Although metadata (information about the digitized work, its provenance and scanning history) is vitally important for scholars wanting to use rely on electronic texts, it is even more important for images, where knowing the quality and reliability of an image is essential. A digital image could come directly from an original object or it could result from a long complex genealogy—derived from one, two or more generations of digital images, themselves taken from transparencies taken from reproductions in publications, for example.

Once technical description and indexing standards are more uniformly accepted, finding related images in different collections on the Internet will be much easier. In the meantime there are interim guideposts on the Web that Trant indicated as good starting points for any image search.

However, a scholar wanting to use digital images will find comparatively little software available for any sophisticated analysis. Leading projects like the Getty-sponsored Museum Educational Site Licensing (MESL) project, bringing museums and educational institutions together, are doing important preliminary work on many levels. One of the many interesting results of this project already appears to be in the experiments with teaching and learning at the college level. These are stimulating new ways of analyzing and using digital images from a variety of sources, and promise to help push the demand for better software.

Trant, now at Archives & Museum Informatics, highlighted the fact that without widely shared standards (technical, data and procedural), the vision of a fully integrated network of digital material will not be realized.

Data

If images are still difficult to work with because of consistency issues, we should expect numbers to be easy. After all, computers were basically built to crunch numbers. Richard Rockwell, director of the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan, surveyed the history of making numerical data available over computer networks. Social science data from charts, from government records, from surveys and experiments have been disseminated broadly but expensively for years. The good news is that many centers are now making the data easily available online and in such flexible forms and with analytical software to enable one to ask a particular question as well as to download pre-packaged data in charts and tables, etc. Rockwell's own ICPSR, the largest social science data archive in the world, has 70% of its material available online.

The problem again is that of integration and documentation. As with texts and images, analyzing the struc-

Building the Scene

Continued from preceding page

ture of a document using SGML, and providing documentation in the form of metadata, is key for any user in navigating material from a wide number of sources.

As the most immediate tasks before us, Rockwell cited the need for a more powerful and more sophisticated Web; a greater involvement by scholars and librarians in creating analytical software; more widespread critical peer-reviews of Web sites and the paramount need for Web Archive projects, so that what is now available on the Web does not get lost.

The AALN

Answering this call for Web site review and furthering Trant's citation of those directories on the Web, Chuck Henry described the innovative American Arts & Letters Network. AALN was created, in what many are calling this interim period of the World Wide Web's dominance of the Internet, to provide a guide to excellence and authenticity on the Web. Sites are reviewed for content and their usefulness for scholars and teachers in a wide range of humanities fields—from aesthetics to theater. Beyond the immediate goal of a quality-assured site, AALN has ambitions to create a collaborative place on the Web for scholars, teachers and students in the humanities.

Commentary

Willard McCarty, a master observer and participant in this field, as editor of "Humanist" and director and teacher of humanities computing, used the panelists' statements to compose a provisional blueprint or "typology of activities for scholarly Internet resources." This "Next Steps" outline pushes forward the political and social ramifications of the panelists' comments. It encourages broad-based integration of the myriad scholarly and cultural projects on the Internet into

Images

Continued from page 8

Trust, 1995), also available at **http://www.gii.getty.edu/intro_imaging/home.html**.

2. The report of this meeting, and other information regarding the development of the Dublin Core and the Warwick Framework for its implementation, can be found at **http://purl.org/metadata/dublin_core**.

3. Available from Cameron Kitchen, American Association for Museums **(cameronk@aol.com)**.

4. A list of the courses taught during the MESL project is available on the Faculty Working Group pages of the MESL WWW site **(http://www.gii.getty.edu/mesl)**. □

the construction of a richer and more relevant medium. His comments are consonant with impulses behind the formation of the National Initiative for a Networked Cultural Heritage. And, like NINCH, he recognizes the importance of engaging with the projects and initiatives in other countries.

Commenting on the presentations, McCarty observed that most of the work lies before us. But he reemphasized the fact that this community is more than fit for the task. Not only have we dreamed of automata since Homer, but the complexity of humanities materials have the potential to push technology's envelope to produce more sophisticated tools. Scholars and scientists need to work hand-in-hand, inspired by the possibilities of mutual benefit. □

Data

Continued from page 10

Statistical Abstract. People will then be able to generate the tables that they need on the fly through Census Bureau Web services.

But what tables about us will then be authoritative? Will people cite a Web site as the source of their evidence? How will they document the procedures they used to generate the tables? How will they and the courts resolve disputes about the evidence?

The private sector will get involved, and different companies will have different ways of producing tables out of the Bureau's Web site. Which company's tables will we believe? Will not society spend more in the aggregate to generate these tables than it would by simply paying the Bureau to do them and to print them?

Further, are government Web sites official records of the United States? If so, they must be archived. But what do you archive? It's hard to do, because you cannot archive a Web site by freezing it for all time! Technological change would soon render it unusable. So when you archive something in this field, you have to continually change it in order to keep it the same. If we don't archive Web sites, then how will the future generation reconstruct what we thought that we knew about ourselves? □

Electronic Texts

Continued from page 5

2. For an overview of electronic text applications in literature, see Susan Hockey, "Electronic Texts in the Humanities: A Coming of Age", p. 21-34 in *Literary Texts in an Electronic Age: Scholarly Implications and Library Services*, edited by Brett Sutton, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign.

3. Renear, Allen, "Representing Text on the Computer: Lessons for and from Philosophy," *Bulletin* of the John Rylands University Library, 74 (1992), 221-248.

4. Eric van Herwijnen, *Practical SGML*, Kluwer, second edition, 1994 is a good introduction to SGML. For a comprehensive review of the SGML world, consult the WWW site at **http://www.sil.org/sgml/sgml.html** compiled and maintained by Robin Cover.

5. Sperberg-McQueen, C.M. and Burnard, L. (eds), *Guidelines for the Encoding and Interchange of Machine-Readable Texts*, TEI document P3, Chicago and Oxford: ACH-ACL-ALLC, 1994. More information is at **http://www.uic.edu/orgs/tei**. See especially the document TEI U5, *TEI Lite: An Introduction to Text Encoding for Interchange* for an introduction to the TEI SGML tag set.

6. See *Optical Character Recognition in the Historical Discipline: Proceedings of an International Workgroup*, Netherlands Historical Data Archive, Nijmegen Institute for Cognition and Information, 1993 for a collection of papers on this topic. Most conclude that OCR is not suitable for humanities source material.

7. The set of metadata elements defined by the first OCLC meeting has become known as the Dublin Core. See Stuart Weibel, Jean Godby, Eric Miller and Ron Daniel, OCLC/NCSA Metadata Workshop Report at **http://www.oclc.org:5046/oclc/research/conferences/metadata/dublin_core_report.html**. (Also see Note 3 of Jennifer Trant's article on digital images.)

8. Notable among these are the Universities of Michigan and Virginia. At the Center for Electronic Texts in the Humanities, we have developed a similar prototype system for the Princeton Library of Electronic Texts in the Humanities (PLETH).

9. The Perseus project (**http://www.perseus.tufts.edu**) is perhaps the best-known humanities hypertext project. It is worth noting that Perseus uses SGML for the archival form of its data.

10. *The Literary Text in the Digital Age*, edited by Richard Finneran (University of Michigan Press), contains a collection of essays on the potential of electronic editions. Peter Robinson's work, most recently with the electronic Wife of Bath's Prologue published on CD-ROM by Cambridge University Press is particularly noteworthy.

11. The MEP is directed by David Chesnutt, editor of the Papers of Henry Laurens. Susan Hockey, Director CETH and C. Michael Sperberg-McQueen, Editor of the TEI serve as coordinators. The partner edition projects are The Documentary History of the First Federal Congress, The Documentary History of the Ratification of the Constitution and the Bill of Rights, The Papers of General Nathanael Greene, The Papers of Henry Laurens, Lincoln Legal Papers, The Papers of Margaret Sanger, and The Papers of Elizabeth Cady Stanton and Susan B. Anthony. □

NINCH

Continued from page 13

Notes

1. *Technology, Scholarship, and the Humanities: The Implications of Electronic Information. September 30-October 1, 1992. Summary of Proceedings* (Getty Trust, 1993)

2. Information Infrastructure Task Force, The National Information Infrastructure: Agenda for Action, 15 September 1993.

3. For more information on the MEP, see Note 11 (above) of Susan Hockey's article on electronic texts. □

AALN

Continued from page 11

Without overly dramatizing the situation, the World Wide Web and the Internet are susceptible to "commodification," becoming defined and priced as a capital resource that does not necessarily allow for universal access or encourage creative and scholarly use. The twentieth century is replete with technologies that began with high expectations of general educational utility, only to be subsequently appropriated for profitable entertainment. Accompanying this kind of transformation (radio and television come first to mind) was a loss of human choice, reducing the engagement of the audience to a rather digital on/off participation.

The stunning increase in networked resources and the speed and capacity of transmission needs to be contextualized at present by the absence of an academic community that could assess these resources, inculcate standards of practice, and delineate benchmarks of success. Fundamental questions concerning the appearance of Web sites in academe—who is the intended audience, why was the site constructed, what are other related sites that might be complementary, how is the usage monitored, and what are the criteria of success—cannot at this time be routinely asked.

One critically important undertaking in the coming years would be to form a community around these digital resources and impose the same kind of scrutiny and evaluation accorded to more traditional printed publications. Without such a community the Web and its considerable potential will most likely drift into a marketplace commodity that may have little value for teaching and research. Conversely, with appropriate scholarly attention, the Web may evolve into a fundamental and indispensable support for higher education, its speed, hyperfunctionality, and multimedia features redefining methodologies as a provocative and catalytic agent. In this respect the future is largely in the hands of the practitioners. □

URLs for Resources Mentioned in this Issue

American Arts & Letters Network (AALN)
http://www.aaln.org

American Council of Learned Societies (ACLS)
http://www.acls.org

American Heritage Virtual Archive Project
http://sunsite.berkeley.edu/FindingAids/EAD/ameriher.html

Archives & Museum Informatics
http://www.archimuse.com

Arts & Humanities Data Service
http://www.kcl.ac.uk/projects/ahds/home.html

ArtServe
http://rubens.anu.edu.au

ArtSource
http://www.uky.edu/Artsource/artsourcehome.html

ArtsWire
http://www.artswire.org

Association for Computers & the Humanities *and*
Association for Literary & Linguistic Computing
http://www.ach.org/

Blake Multimedia Project
http://www.fmdc.calpoly.edu/libarts/smarx/Blake/blakeproject.html

Brown University Women Writers' Project
http://www.stg.brown.edu/projects/wwp/wwp_home.html

CH Working Papers
http://www.chass.utoronto.ca/epc/chwp/ *and*
http://www.kcl.ac.uk/kis/schools/hums/ruhc/chwp/

Christus Rex
http://www.christusrex.org

The Complete Writings & Pictures of Dante Gabriel Rossetti
http://jefferson.village.virginia.edu/rossetti/rossetti.html

Data Centre, Computing in the Humanities & Social Sciences, University of Toronto
http://datacentre.chass.utoronto.ca/

Data Documentation Initiative
http://www.icpsr.umich.edu/DDI

Dublin Core
http://purl.org/metadata/dublin_core *and*
http://www.oclc.org:5046/oclc/research/conferences/metadata/dublin_core_report.html

fineArt forum
http://www.msstate.edu/Fineart_Online/home.html

Health and Retirement Survey, University of Michigan
http://www.umich.edu/~hrswww/

Humanist
http://www.princeton.edu/~mccarty/humanist/

Inter-university Consortium for Political & Social Research (ICPSR)
http://www.icpsr.umich.edu/

Introduction to Imaging: Issues in Constructing an Image Database
http://www.gii.getty.edu/intro_imaging/home.html

Labyrinth, Georgetown University
http://www.georgetown.edu/labyrinth

Model Editions Partnership
http://mep.cla.sc.edu

Multimedia Yeats
http://www.oit.itd.umich.edu/WIP/ArtsHumanities/MultimediaYeats.html

Museum Educational Site Licensing Project (MESL)
http://www.gii.getty.edu/mesl

National Digital Library Federation
http://lcweb.loc.gov/loc/ndlf

National Digital Library Program
http://lcweb2.loc.gov/ammem/ammemhome.html

National Initiative for a Networked Cultural Heritage (NINCH)
http://www-ninch.cni.org

National Museum of American Art
http://www.nmaa.si.edu

Orlando Project
http://www.ualberta.ca/ORLANDO

Panel Study of Income Dynamics, University of Michigan
http://www.umich.edu/~psid/

Perseus Project
http://www.perseus.tufts.edu

Piero Project
http://mondrian.princeton.edu/art430/art430.html

Pompeii Forum Project
http://jefferson.village.virginia.edu/pompeii/page-1.html

SGML Web Page
http://www.sil.org/sgml/sgml.html

Text Encoding Initiative
http://www.uic.edu/orgs/tei

Texte: Revue de Critique et de Théorie Littéraire
http://www.chass.utoronto.ca:8080/french/litera/Revue_Texte/Texte.html

U.S. Bureau of the Census
http://www.census.gov/

Vergil Project
http://ccat.sas.upenn.edu/~joel/courses/project.html

Virtual Library Virtual Museum
http://www.icom.org

WebMuseum, Paris
http://sunsite.unc.edu/wm/

Paul Evan Peters (1947-1996)

Paul Evan Peters was a partner with the ACLS in the conception and establishment of the National Initiative for a Networked Cultural Heritage. We mourn his loss. The following is excerpted from the *Association of Research Libraries Newsletter* 189.

Paul Evan Peters, a charismatic leader in the development and expanded use of electronically networked information by the scholarly community, died suddenly on November 18, 1996, while walking on a Florida beach with his wife during a visit to his parents' home. He was 48. The cause of death was an asthma attack.

Paul was the founding director of the Coalition for Networked Information (CNI), which was formed in 1990 to promote the creation and use of networked information resources and services that advance scholarship and intellectual productivity. In the six years that he led the Coalition, he shaped it into a highly regarded and influential forum where the higher education (including virtually all major research universities), library, information science, and scholarly communities could effectively engage the challenge of networked information.

He sought common ground for many constituencies to develop global networked information resources. His sudden death stunned the communities where he was most admired and where his vision and ability to pull people together to utilize the new realities of the Internet were seen as unique.

Before founding the Coalition in March 1990, Paul was systems coordinator at the New York Public Library from 1987 through 1989, and was Assistant University Librarian for Systems at Columbia University, where he also earned a master's degree in sociology in 1986. From 1970 until 1978, he was principal in a variety of research and development projects and earned a master's degree in library and information science at the University of Pittsburgh. He earned undergraduate degrees in computer science and philosophy at the University of Dayton in 1969. Paul was a former president of the Library and Information Technology Association, was a former chair of the National Information Standards Organization, and served on the editorial board of a number of networking, networked information, and library technology journals.

Some of Paul's most outstanding accomplishments also included: his service on the council of the American Library Association; his position as curator for The Global Library Exhibit at the New York Public Library; and his work to develop the National Initiative for a Networked Cultural Heritage (NINCH). He was also a catalyst for electronic journal, theses, and archives projects as well as mentoring and providing vision to the library and information technology community. In short, Paul was an imagineer.

Paul was born on December 12, 1947, in Dayton, Ohio. He is survived by his wife Rosemarie Kozdron, his parents Austin and Mary Peters, and a brother Philip.

As a tribute to Paul's commitment, the Coalition Task Force meeting was held as planned December 6-7, 1996, in San Francisco. Over 320 representatives from the communities he served gathered to engage the agenda Paul had planned for the meeting—enterprise-wide information strategies. A memorial service was held during the Task Force Meeting and a scholarship fund was announced.

A memorial service in Washington, DC will be held on February 18, 1997, at 4 p.m. at Georgetown University's Dahlgren Chapel. A reception will follow.

Further information is available at **http://www.cni.org/prog-oview.html**.

American Council of Learned Societies
228 East 45th Street, New York, NY 10017-3398

Non-Profit
Organization
U.S. Postage
PAID
Moorestown, NJ
Permit No. 15

ACLS

Donald P Ely
Syracuse Univ Sch of Ed
Eric Clearinghouse
Info Res
Syracuse NY 13210